Performance Analysis on Cache-Enabled FR2 IAB Networks

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 6, 2021, accepted April 1, 2021, date of publication April 6, 2021, date of current version May 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071355

Performance Analysis on Cache-Enabled FR2 IAB Networks

TONG ZHANG[1], SUDIP BISWAS[2], (Member, IEEE), AND THARMALINGAM RATNARAJAH[1], (Senior Member, IEEE)

[1]School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh EH9 3FG, U.K.
[2]Electronics and Communications Engineering Department, Indian Institute of Information Technology Guwahati, Guwahati 781015, India

Corresponding author: Tong Zhang (t.zhang@ed.ac.uk)

This work was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P009549/1, in part by the U.K.–India Education and Research Initiative Thematic Partnerships under Grant DST UKIERI-2016-17-0060, and in part by the Scheme for Promotion of Academic and Research Collaboration (SPARC) Project 148. The work of Sudip Biswas was supported by the Science and Engineering Research Board (SERB), Government of India, under Grant SRG/2020/001145.

ABSTRACT This paper brings the concept of wireless edge caching (WEC) to the realm of integrated access and backhaul (IAB) networks. Traditionally, both access and backhaul are allocated fixed spectrum. However, when WEC is applied to the IAB nodes to pre-fetch some popular files, the backhaul traffic can be drastically reduced. Accordingly, an analytical framework is developed for a cache-enabled IAB network for the wide-band frequency range 2 (FR2) ($>$ 24.25 GHz) channel model. In particular, the optimal spectrum resource allocation policy between access and backhaul for IAB networks is investigated by utilizing WEC. For the considered WEC enabled IAB network, a mathematical framework is presented under the content delivery phase for the calculation of three key performance metrics, namely the average success probability (ASP), throughput, and latency of file delivery with respect to various network parameters such as different caching placement strategies, node density, cache size, antenna number, blockage density, and resource partitioning factor. Extensive numerical results are provided to both verify the theoretical derivations and to demonstrate the precedency of using WEC as a cost-effective measure for the improvement of the IAB network's performance. In particular, under a normalized cache size of 0.6, compared to the baseline no caching scenario, the percentages of backhaul spectrum that can be saved are 48% (26%), 66% (48%), and 47% (27%) when caching the most popular files (caching uniformly) with respect to the ASP, latency, and throughput of file delivery, respectively. The saved backhaul spectrum can then be shifted to the access for its performance enhancements. Our findings also show that there exists an optimal spectrum partition for IAB with respect to different network parameter settings. Further, there exists a tradeoff in the selection of the optimal partitioning factor with respect to ASP/ throughput and latency of file delivery for varying IAB node densities.

INDEX TERMS Wireless edge caching, 5G NR, stochastic geometry, integrated access and backhaul networks, hybrid beamforming, spectrum partitioning.

I. INTRODUCTION

In order to meet the exponential rise in demand for wireless data services, the fifth generation (5G) and beyond cellular networks consider the frequency range 2 (FR2) spectrum, which is the milimeter wave (mmWave) bands greater than 24.25 GHz due to the availability of abundant unused frequency bands. Further, to mitigate the effects of path loss and blockages that plague transmission in FR2 bands, network densification is considered in literature. Nevertheless, the performance gains achieved using FR2 transmission and network densification are highly dependent on the high-capacity backhaul links since the total data traffic from the access links is inevitably forwarded to the core network through the backhaul networks. In particular, as the number of small base stations (SBSs) increases, the traffic intensity during peak hours of the backhaul link significantly increases, which is the primary bottleneck of emerging dense wireless networks. In this regard, 3GPP NR release–16 proposed the novel concept of integrated access and backhaul (IAB) architecture

The associate editor coordinating the review of this manuscript and approving it for publication was Yunlong Cai.

that allows the IAB nodes to share the same FR2 spectrum between access and backhaul resources. The highly directional beamforming applicable to FR2 makes it possible to replace the so called last-mile fibers for SBSs by establishing mmWave backhaul links between the SBS and the macro base station (MBS) equipped with fiber backhaul. Accordingly, the IAB node provides wireless access to users, while also receiving backhaul wirelessly from IAB donor nodes under the same FR2 spectrum resource with either in-band operation (*i.e.,* the access link operates in the same frequency band as the backhaul link) or out-of-band operation (*i.e.,* the access and backhaul links operate in separate frequency bands) depending on the data requirement trend of the network. Once implemented, network operators will be able to quickly add new IAB nodes dynamically without the constraint of wired backhaul.

However, designing an efficient and high-performance IAB network is a challenging communication engineering problem, which necessitate radically novel communication concepts in the physical layer. For example, the authors in [1] analyzed the coverage performance of an IAB network with respect to various backhaul spectrum partition policies and demonstrated that the dynamic FR2 spectrum partition between access and backhaul links is a technological problem.

On a similar note, wireless edge caching (WEC) is another promising solution to effectively overcome the backhaul bottleneck since a large amount of the resultant traffic is caused by redundant transmissions of some popular files [2]. By pre-fetching these popular files in the local caches of the IAB nodes, on one hand, the backhaul load can be effectively minimized, while on the other hand, the content can be brought closer to users at the node edge, which shortens the transmission distance, thus reducing end-to-end latency. Recently, WEC has gained significant attention with many undergoing studies on the design and analysis of cache-enabled cellular networks. Accordingly, the optimal caching placement to minimize the expected downloading delay in the femto-caching networks was proposed in [3], while the authors in [4] analyzed the optimal caching placement algorithm with respect to minimizing the bit error rate for wireless femto-caching networks. Further, stochastic-geometry-based cache-enabled small cell networks was analyzed in terms of the outage probability and average content delivery rate in [5] and cache miss probability with respect to storage-bandwidth tradeoff in [6]. Next, 5G enabling solutions amalgamated with WEC, such as cache-enabled FR2 networks and massive multi-input multi-output (mMIMO)-aided self-backhauled dense HetNets were studied in [7] and [8]–[10], respectively. Further, WEC was applied to FR1-FR2 hybrid networks in [11], where the optimal caching placement algorithm was proposed to minimize the average success probability (ASP) of file delivery.

Against this background, this paper investigates an innovative spectrum resource allocation solution for IAB through the incorporation of WEC. By implementing caching, the users whose requested files are available at local caches of IAB nodes can be directly served through an one-hop access link, which can effectively 1) reduce the backhaul loads, 2) reduce the backhaul requirement (*e.g.,* less cache-miss users in access will require less throughput requirement), and 3) save backhaul resource. Consequently, WEC provides the opportunity to shift the additional saved backhaul resource to the access link in pursuit of better performance. The primary distinctions of this work are summarized as follows:

- We consider the wide-band FR2 channel model that takes the frequency selectivity into account. Further, in order to eliminate the inter-symbol interference (ISI) and effectively use spatial multiplexing, we consider that each IAB node (IAB donor node) simultaneously serves multiple users in access (users in access and IAB nodes in backhaul) through fully-connected OFDM-based hybrid beamforming architecture. In particular, since FR2 communication is more likely to be noise-limited in high blockage environments, we characterize the signal-to-noise-ratio (SNR) and the instantaneous rate for both access and backhaul with respect to cache-hit and cache-miss scenarios under the out-of-band dynamic resource allocation policy.
- Next, we resort to stochastic geometry tools to model the stochastic locations of cache-enabled IAB nodes, IAB donor nodes, and users, with which we first derive the association probabilities for users in access and IAB nodes with respect to LOS and NLOS transmissions, the probability density functions (PDFs) of the serving distance, and approximated cell loads. In particular, in order to characterize the effect of caching on backhaul requirement and resource allocation strategy, we assume that each user has a fixed target rate and accordingly use the approximated cell loads to derive the average target rate per link to derive and evaluate the ASP, throughput, and latency of file delivery with respect to node density, cache size, blockage density, antenna number, and bandwidth partitioning factor.
- Finally, through extensive numerical results, we demonstrate the efficacy of caching in reducing the core network signalling load. In particular, we show that caching can effectively improve the performance of an end-to-end IAB network involving both backhaul and access when compared to a traditional IAB network without caching. We show that larger cache size leads to a greater shift of bandwidth resources from backhaul to the access links, while more blockages leads to less bandwidth shifts. Also, our results demonstrate that there exists an optimal spectrum resource allocation factor for access and backhaul that provides the maximum ASP, throughput, and latency of file delivery under different parameter settings. Further, we also show the gap between the best-case scenario (i.e., fully cached case) and a general case. We also demonstrate that increasing the antenna number while choosing a proper partitioning

TABLE 1. Summary of notations.

Notations	Physical meaning
Φ_m, Φ_s, Φ_u	PPP-based locations of IAB donor nodes, IAB nodes, and users, respectively
$\lambda_m, \lambda_s, \lambda_u$	Spatial densities of IAB donor nodes, IAB nodes, and users, respectively
$\mathrm{P}_m, \mathrm{P}_s$	Transmit power of each IAB donor node and IAB nodes, respectively
n_t^m, n_t^s	Number of transmit antennas at each IAB donor node, IAB node, respectively
n_r^s, n_r^u	Number of receive antennas at each IAB node and user, respectively
$n_{\mathrm{RF}}^s, n_{\mathrm{RF}}^m$	Number of RF chains at each IAB node and IAB donor node, respectively
W_{bw}	Total bandwidth
$\mathcal{F} = \{f_1, \ldots, f_F\}$ (*i.e.*, $\lvert\mathcal{F}\rvert = F$)	The finite file set with F total files
S	File size in bits
$\mathcal{Q} = \{q_1, \ldots, q_F\}$	Content popularity vector with q_i denoting the probability of request of the ith file ($i \in \{1, \ldots, F\}$)
υ	Skewness of Zipf distribution
C	Cache size of each IAB node
$\tilde{\Omega} = \{\tilde{\omega}_1, \ldots, \tilde{\omega}_F\}$	Caching probability set with $\tilde{\omega}_i$ denoting the probability that the file i is cached in the IAB node
$\hat{m}_{\mathcal{L}}, \hat{m}_{\mathcal{N}}$	Nakagami parameters for LOS and NLOS paths
M_m^u (M_m^s), M_s^u	Maximum number of users (IAB nodes) served by each IAB donor node and users served by IAB node
N_x^{mu} (N_x^{ms}), N_x^s	Number of associated users (IAB nodes) for each IAB node and users served by IAB donor node (located at x)
$\mathcal{U}_x^m, \mathcal{U}_x^s$	Set of scheduled IAB nodes plus users served by each IAB donor node and users served by IAB node located at x
U_x^m, U_x^s	Cardinality of the set $\mathcal{U}_x^m$ and $\mathcal{U}_x^s$, respectively
β	Blockage density
$\alpha_{\mathcal{L}}, \alpha_{\mathcal{N}}$	LOS and NLOS path loss exponent, respectively
d	Time delay of each path
$\eta_{\mathcal{L}}, \eta_{\mathcal{N}}$	Number of LOS and NLOS scatters, respectively
$\mathbf{a}_R(\cdot), \mathbf{a}_T(\cdot)$	Steering vector responses of the receiver and transmitter, respectively
θ, ϕ	Angles of arrival and departure, respectively
T_s	Sampling time
ϵ	The roll-off factor of the raised-cosine filter
ν_u	Target rate per user
D	Cyclic prefix
K	Total number of subcarriers
$\mathcal{P}_s(\nu), \bar{R}, \bar{T}_{(\cdot)}$	Average success probability, throughput, and latency of file delivery, respectively
σ^2	Noise power
N_{bw}	Partitioning factor

factor is an effective method to improve the performance of the network, that can even surpass the performance achieved under the best-case scenario. Finally, we show the tradeoff in selecting the optimal partitioning factor between the ASP/throughput and latency performance under different IAB node density settings.

II. SYSTEM MODEL

In this section, we describe the considered WEC-enabled FR2 IAB network model. The main notations used in the paper are summarized in Table 1.

A. NETWORK TOPOLOGY

We consider a storage equipped FR2 IAB network as shown in Fig. 1, that consists of IAB donor nodes with transmit power P_m and cache-enabled IAB nodes with transmit power P_s, with $\mathrm{P}_s < \mathrm{P}_m$, which are distributed in the 2-D Euclidean space $\mathbb{R}^2$ according to independent homogeneous Poisson point processes[1] (PPPs) Φ_m and Φ_s with densities λ_m and λ_s, respectively. The locations of users follow another independent PPP Φ_u with density λ_u. The link from IAB donor node/IAB node to the user is referred to as the access link while the link from IAB donor node to the IAB node

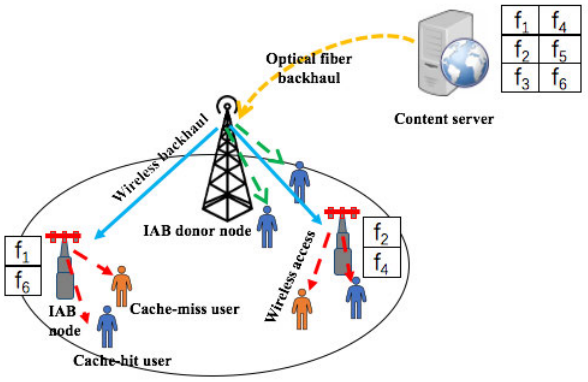

FIGURE 1. An illustration of a cache-enabled IAB FR2 network.

is referred to as the backhaul link. According to Slivnyak's theorem [12], which states that the statistical properties of all location-dependent random parameters seen at a random point of a PPP Φ is identical to those seen at the origin in the process $\Phi \cup \{0\}$, without any loss of generality, the analysis hereinafter is performed for the typical user located at the origin. The IAB donor nodes are assumed to be connected to the core network via high-capacity optical fiber, such that the focus of the work lies in the analysis of wireless backhaul between IAB nodes and IAB donor nodes.

[1]In the field of probability and statistics, a Poisson point process is defined as a type of random mathematical object consisting of points that are randomly located on a mathematical space [12]–[14].

Further, we assume that all files are available in the IAB donor nodes [15] while IAB nodes are equipped with limited storage memory to cache certain popular files. Thus, if the user is served by either the IAB donor node or the IAB node that stores the requested files in the local caches, the user will experience only one-hop access link. However, if the user is served by the IAB node that does not store the requested files in the local caches, the IAB node first retrieves the non-cached files from the IAB donor node through the backhaul link and then serves the user through the access link, whereby the user will experience a two-hop link. The propagation between each IAB node and its associated users in access links takes place via fully connected hybrid precoders that combine both radio frequency (RF) and baseband (BB) precoders. Similarly, each IAB donor node simultaneously serves its associated users in access and its associated IAB nodes (*i.e.,* the cache miss users in access links of the associated IAB nodes) in backhaul[2] through the same hybrid beamforming architecture to achieve high data rates.

Next, due to very short wavelength of FR2 signals, it is reasonable to assume that even small size of the mobile phones can accomodate multiple antennas [11]. In addition, because of the sparsity of the FR2 signal propagation characteristic, all scatters are assumed to take place in the azimuth plane and are uniformly distributed in $[0, 2\pi]$ [16]. Therefore, we assume that each IAB donor node, IAB node, and user is equipped with uniform linear array (ULA) of size n_t^m, n_t^s, and n_r^u, respectively [16], [17]. In particular, each IAB donor node and IAB node has n_{RF}^m and n_{RF}^s RF chains, respectively, for transmitting. Besides, since the main interest of the work lies in focusing on the effects of caching and spectrum resource partitioning for IAB networks, for analytical tractability we assume that multiple users (or IAB nodes) are served by its serving IAB node or IAB donor node in access (or IAB donor node in backhaul) through a single stream[3] per user (or IAB node). Thus, it is sufficient to assume that each user (or IAB node) employs a RF-only combiner with the number of antennas n_r^u (or n_r^s) to denote the received signals as verified in [16]. More details on the propagation model are presented in the subsequent parts of this section.

B. CONTENT POPULARITY AND CACHING MODEL

The file set is denoted by $\mathcal{F}$, whereby there are a total of $|\mathcal{F}| = F$ files in the file set such that $\mathcal{F} = \{f_1, \ldots, f_F\}$. Each file is assumed to be of equal file size[4] (S bits). Each user makes independent requests for files from the file set $\mathcal{F}$ with a requesting probability that follows the Zipf distribution, which is a commonly used distribution for video and web pages popularity [11], [18]–[20]. Further, we define the request probability set as $\mathcal{Q} = \{q_1, \ldots, q_F\}$, with the probability of requesting the ith file given as $q_i = (i^{\upsilon} \sum_{j=1}^{F} j^{-\upsilon})^{-1}$, where υ is the Zipf skewness parameter that tunes the content popularity distribution. Since this work's interest mainly focuses on investigating the impact of spectrum partitioning from caching allocation perspective, we assume that the content popularity is known in prior.[5]

Assume that the cache size of each IAB node is C. During off-peak hours, content caching placement is performed for the cache-enabled IAB nodes based on certain caching placement strategies. This work considers a generic caching placement policy in a probabilistic manner so that it makes the analysis of different caching placement policies in the stochastic geometric framework tractable. Further, we define a caching probability set $\tilde{\Omega} = \{\tilde{\omega}_1, \ldots, \tilde{\omega}_F\}$, such that $0 \leq \tilde{\omega}_i \leq 1$ with $\forall i \in [1, F]$ according to the property of probability and $\sum_{i=1}^{F} \tilde{\omega}_i \leq C$ based on the cache size constraint in an average sense. Specifically, this work analyzes two commonly used caching placement strategies: 1) caching most popular files (MC) and 2) uniform caching (UC). As for MC, the caching probability $\tilde{\omega}_i = 1$ for $\forall i \in [1, C]$ and $\tilde{\omega}_i = 0$ for other files. On the other hand, instead of caching C most popular files, UC considers that all files are stored in the local caches with the caching probability $\tilde{\omega}_i = \frac{C}{F}$ for $\forall i \in [1, F]$. Accordingly, the cache hit ratio of the IAB node is given as

$$p_{\text{hit}} = \sum_{i=1}^{F} q_i \tilde{\omega}_i = \frac{\sum_{i=1}^{F} i^{-\upsilon} \tilde{\omega}_i}{\sum_{j=1}^{F} j^{-\upsilon}}. \tag{1}$$

Finally, we also consider the scenario without caching (NC) as a baseline for comparison, where the IAB nodes are not equipped with any storage memories and cannot cache any files.

Remark 1: MC and UC are the most commonly used proactive caching placement schemes that are primarily influenced by the content popularity distribution. In particular, caching the most popular files is optimal for maximizing the throughput, while uniform caching provides the maximum file diversity and is preferable for maximizing the success probability. Although there are several other caching schemes and algorithms in literature, most of them are application/ model specific and might not be suitable when other models (the IAB network in this case) are considered.

[2]We assume that the files sent to the cache miss users of the IAB node by the IAB donor node through backhaul are first encoded together as one file and then simultaneously delivered along with the files requested by users of the IAB donor node in its access links. After receiving the coded file at the IAB node, the received non-cached coded file is decoded as an independent requested file and then forwarded to users in access. In particular, the IAB node will not require the backhaul if and only if there is no cache miss users associated with the IAB node. However, the quality of experience of users on backhaul is highly related to the cache miss users since more cache miss users require more resources on backhaul links, which will be discussed later.

[3]The multi-stream scenario with multiple RF chains at the receiver is also applicable for the presented analytical framework, whereby the total rate should be computed by multiplying the number of receive RF chains with the data rate of the single stream.

[4]Unequal file sizes can be partitioned into equal size small chunks with any loss of generality. Thus, the analytic framework of this work is justifiable with the assumption of equal size.

[5]The content popularity can be estimated by using prediction algorithms [21], [22], which is beyond the scope of this work.

C. DOWNLINK mmWave PROPAGATION MODEL

1) BLOCKAGE MODEL

Unlike traditional FR1 signals, FR2 signals with short wave-length are susceptible to blockages in the network, such as concrete walls and trees, etc., which can result in either direct line-of-sight (LOS) paths or reflected non-LOS (NLOS) blocked propagation paths. Therefore, we consider stationary blockages that are invariant to direction and thus, the blockage effects in this work are modelled based on a two-state stationary probabilistic "exponential blockage model" proposed and validated in [23], where the probability of the propagation path being LOS and NLOS states are given as $p_{\mathcal{L}}(r) = e^{-\beta r}$ and $p_{\mathcal{N}}(r) = 1 - p_{\mathcal{L}}(r)$, respectively, with β referred to as the blockage density and r denoting the link distance between the transmitter and receiver. Further, based on the distance-dependent LOS/NLOS probability functions, the original PPP-based nodes are further thinned into two independent non-homogeneous point processes on the plane given as $\Phi_t^{\mathcal{L}}$ and $\Phi_t^{\mathcal{N}}$ with densities $\lambda_t p_{\mathcal{L}}(r)$ and $\lambda_t p_{\mathcal{N}}(r)$, respectively, where $t \in \{m, s\}$ denotes either the set of IAB donor nodes or IAB nodes. Accordingly, with respect to the blockage effects, the path loss exponent on each link is a random variable given as

$$\alpha = \begin{cases} \alpha_{\mathcal{L}} & \text{w.p.} \quad p_{\mathcal{L}}(r) \\ \alpha_{\mathcal{N}} & \text{w.p.} \quad p_{\mathcal{N}}(r) \end{cases}, \tag{2}$$

where $\alpha_{\mathcal{L}}$ and $\alpha_{\mathcal{N}}$ are the LOS and NLOS path loss exponents, respectively.

2) CHANNEL MODEL

Due to high free-space path loss, the FR2 propagation environment is well characterized by a clustered channel model. Different from the narrowband assumption, where the time delays of different clusters are small compared to the sampling period and the time delays are ignored [16], this work considers the wide-band FR2 channel model to incorporate the frequency selective effects. Further, each path is modelled by a time delay d, in addition to the complex gain as well as the angles of departure and arrival [24], [25]. Therefore, the delay-d MIMO channel matrix between the transmitter at x and the receiver at y based on the geometric channel model is given as

$$\tilde{\mathbf{H}}_{xy}[d] = \sqrt{\frac{n_t n_r}{r_{xy}^{\alpha} \eta_{xy}}} \sum_{l=1}^{\eta_{xy}} h_{lxy} p_{\text{rc}}(dT_s - \tau_l) \mathbf{a}_R(\theta_{lxy}) \mathbf{a}_T^H(\phi_{lxy}), \tag{3}$$

where n_r and n_t are receive and transmit antennas, respectively. h_{lxy} is the complex small-scale fading gain of the lth cluster. This work assumes that $|h_{lxy}|^2$ follows independent Nakagami distribution for each link such that the channel gain for the signal $|h_{lxy}|^2 \sim \Gamma(\hat{m}, \frac{1}{\hat{m}})$, where $\hat{m} \in \{\hat{m}_{\mathcal{L}}, \hat{m}_{\mathcal{N}}\}$ is the Nakagami fading parameter depending on the LOS and NLOS. $\alpha \in \{\alpha_{\mathcal{L}}, \alpha_{\mathcal{N}}\}$ is the path loss exponent and $\eta_{xy} \in \{\eta_{\mathcal{L}}, \eta_{\mathcal{N}}\}$ is the number of scatters depending on LOS or NLOS path such that $\eta_{\mathcal{L}} < \eta_{\mathcal{N}}$ [16]. In particular, we denote the angles of departure and arrival as ϕ and θ, respectively. Subsequently, the steering vectors of the transmitter (*i.e.,* IAB nodes for $T = s$ or IAB donor nodes for $T = m$) and receiver (*i.e.,* users for $R = u$ or IAB nodes for $R = s$) are denoted as $\mathbf{a}_T(\phi)$ and $\mathbf{a}_R(\theta)$, respectively, which are given as

$$\mathbf{a}_T(\phi) = \frac{1}{\sqrt{n_t}} \left[1 \; e^{j\tilde{k}\sin(\phi)} \; \ldots \; e^{(n_t-1)j\tilde{k}\sin(\phi)}\right]^T, \tag{4}$$

$$\mathbf{a}_R(\theta) = \frac{1}{\sqrt{n_r}} \left[1 \; e^{j\tilde{k}\sin(\theta)} \; \ldots \; e^{(n_r-1)j\tilde{k}\sin(\theta)}\right]^T. \tag{5}$$

Here, $\tilde{k} = \frac{2\pi\tilde{d}}{\lambda}$, with λ denoting the wavelength and $\tilde{d}$ refers to the distance between antenna elements. $p_{\text{rc}}(\tau)$ denotes the pulse shaping function for T_s-spaced signaling evaluated at τ seconds, where $p_{\text{rc}}(t)$ is given as

$$p_{\text{rc}}(t) = \begin{cases} \frac{\pi}{4}\text{sinc}(\frac{1}{2\epsilon}), & t = \pm\frac{T_s}{2\epsilon} \\ \text{sinc}(\frac{t}{T_s})\frac{\cos(\frac{\pi\epsilon t}{T_s})}{1-(\frac{2\epsilon t}{T_s})^2}, & \text{otherwise} \end{cases}, \tag{6}$$

where T_s is the sampling time and $\epsilon = 1$ is the roll-off factor.

Further, this work considers OFDM transmission and assume the total number of subcarriers be K and the length of cyclic prefix be D. Given the delay-d MIMO channel matrix, the channel at the subcarrier k, denoted as $\mathbf{H}_{xy}[k]$, can be characterized with the assumption of the perfect synchronization as

$$\mathbf{H}_{xy}[k] = \sum_{d=0}^{D-1} \bar{\mathbf{H}}_{xy}[d] e^{-j\frac{2\pi k}{K}d}, \tag{7}$$

For notational simplicity, the channel matrix at the kth subcarrier is given as

$$\mathbf{H}_{xy}[k] = \sqrt{\frac{n_t n_r}{r_{xy}^{\alpha} \eta_{xy}}} \sum_{l=1}^{\eta_{xy}} h_{lxy} \omega_{\tau_l}[k] \mathbf{a}_R(\theta_{lxy}) \mathbf{a}_T^H(\phi_{lxy}), \tag{8}$$

where $\omega_{\tau_l}[k] = \sum_{d=0}^{D-1} p_{\text{rc}}(dT_s - \tau_l) e^{-j\frac{2\pi k}{K}d}$. In particular, the path delay is uniformly distributed in $[0, DT_s]$.

III. USER ASSOCIATION AND CELL LOAD

In this section, we derive the association probabilities, the PDFs of the serving distance, and the cell load according to the least biased path loss association strategy. Further, for analytical tractability, we consider the typical backhaul link similar to the access link with the relevant receiver (*i.e.,* the tagged IAB node that is served by the IAB donor

TABLE 2. Association events.

Event	Description
$E_{am}^{\mathcal{L}}$	The typical user is associated with the IAB donor node in LOS path located at $x_m^{\mathcal{L}} \in \Phi_m^{\mathcal{L}}$
$E_{am}^{\mathcal{N}}$	The typical user is associated with the IAB donor node in NLOS path located at $x_m^{\mathcal{N}} \in \Phi_m^{\mathcal{N}}$
$E_{as}^{\mathcal{L}}$	The typical user is associated with the IAB node in LOS path located at $x_s^{\mathcal{L}} \in \Phi_s^{\mathcal{L}}$
$E_{as}^{\mathcal{N}}$	The typical user is associated with the IAB node in NLOS path located at $x_s^{\mathcal{N}} \in \Phi_s^{\mathcal{N}}$
$E_{bm}^{\mathcal{L}}$	The tagged IAB node is associated with the IAB donor node in LOS path located at $y_m^{\mathcal{L}} \in \Phi_m^{\mathcal{L}}$
$E_{bm}^{\mathcal{N}}$	The tagged IAB node is associated with the IAB donor node in NLOS path located at $y_m^{\mathcal{N}} \in \Phi_m^{\mathcal{N}}$

node serves the typical user) located at the origin[6] to ignore the correlation of locations between the serving IAB donor node and the tagged IAB node without compromising on the accuracy of the results and design insights [26], [27]. In addition, by means of this assumption, we can manipulate the cell load in an approximated form that will be given in the following subsection. Hereinafter, we denote the typical receiver (*i.e.,* either the typical user or tagged IAB node) by the subscript 0. Further, let the location of the IAB node that serves the typical user in access link be denoted as $x_t \in \Phi_t$, where $t = m$ is referred to as the IAB donor node and $t = s$ is referred to as the IAB node. Therefore, the user association according to the least biased path loss is formulated by $\max_{x_t \in \Phi_t} B_t r_{x_t,0}^{-\alpha}$, where B_t is the bias factor to control the extension or shrinkage of the coverage and $r_{x_t 0}$ is the distance between the transmitter and the receiver. Further, to achieve more offloading from the IAB donor nodes, we have $B_s \geq B_m$. Similarly, let the location of the IAB donor node that serves the tagged IAB node in backhaul link be denoted as $y_m \in \Phi_m$, such that the IAB node association according to the least biased path loss is reduced to $\max_{y_m \in \Phi_m} r_{y_m 0}^{-\alpha}$.

A. ASSOCIATION PROBABILITIES

Due to LOS and NLOS transmissions, the coverage area of IAB donor nodes and IAB nodes no longer follow the weighted Voronoi cells since the typical user (the tagged IAB node) can associate to a far away IAB donor/IAB node (IAB donor node) with LOS path rather than a nearby one with NLOS path. Therefore, we define the association events as shown in the Table 2. Subsequently, the association probabilities related to the relevant events are defined as $\mathcal{A}_{am}^{\mathcal{L}} = \mathbb{P}[E_{am}^{\mathcal{L}}]$, $\mathcal{A}_{am}^{\mathcal{N}} = \mathbb{P}[E_{am}^{\mathcal{N}}]$, $\mathcal{A}_{as}^{\mathcal{L}} = \mathbb{P}[E_{as}^{\mathcal{L}}]$, $\mathcal{A}_{as}^{\mathcal{N}} = \mathbb{P}[E_{as}^{\mathcal{N}}]$, $\mathcal{A}_{bm}^{\mathcal{L}} = \mathbb{P}[E_{bm}^{\mathcal{L}}]$, and $\mathcal{A}_{bm}^{\mathcal{N}} = \mathbb{P}[E_{bm}^{\mathcal{N}}]$. The following proposition provides the semi-closed form expressions of the association probabilities of these events.

Proposition 1: The association probabilities of the typical user associated with the IAB node in LOS and NLOS are respectively given as

$$\mathcal{A}_{as}^{\mathcal{L}} = \int_0^{\infty} \exp\Bigg[-2\pi\lambda_s \left(\hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) + Z\left(R\right)\right) - 2\pi\lambda_m \left(\hat{Z}\left(\left(K_B R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right) + Z\left(\left(K_B R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right)\right)\Bigg] 2\pi\lambda_s Z'(R) dR, \tag{9}$$

$$\mathcal{A}_{as}^{\mathcal{N}} = \int_0^{\infty} \exp\Bigg[-2\pi\lambda_s \left(\hat{Z}\left(R\right) + Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right)\right) - 2\pi\lambda_m \left(\hat{Z}\left(\left(K_B R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right)\Bigg] 2\pi\lambda_s \hat{Z}'(R) dR, \tag{10}$$

where $Z'(R) = Re^{-\beta R}$, $\hat{Z}'(R) = R - Re^{-\beta R}$, *and* $K_B = \frac{B_m}{B_s}$. *And the association probabilities of the typical user associated with the IAB donor node in LOS and NLOS are respectively given as*

$$\mathcal{A}_{am}^{\mathcal{L}} = \int_0^{\infty} \exp\Bigg[-2\pi\lambda_s \left(\hat{Z}\left(\left(K_B^{-1} R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B^{-1} R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right) - 2\pi\lambda_m \left(\hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) + Z\left(R\right)\right)\Bigg] 2\pi\lambda_m Z'(R) dR, \tag{11}$$

$$\mathcal{A}_{am}^{\mathcal{N}} = \int_0^{\infty} \exp\Bigg[-2\pi\lambda_s \left(\hat{Z}\left(\left(K_B^{-1} R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B^{-1} R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right) - 2\pi\lambda_m \left(\hat{Z}\left(R\right) + Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right)\right)\Bigg] 2\pi\lambda_m \hat{Z}'(R) dR. \tag{12}$$

Similarly, the association probabilities of the tagged IAB node associated with the IAB donor node in LOS and NLOS

[6]In fact, the original PPP of IAB donor nodes formed according to the reference point (*i.e.,* the typical user) is no longer the exactly same homogeneous PPP for the receiver (*i.e.,* the tagged IAB node) since the location of the serving IAB donor node is conditioned on the location of the tagged IAB node. Put another way, given the typical user is served by the tagged IAB node, only the set of IAB donor nodes that cannot serve the typical user are considered as the potential serving IAB donor nodes for the tagged IAB node; after that, the density of the potential serving IAB donor nodes is computed by moving the tagged IAB node to the origin, which requires the density of the potential serving IAB donor nodes related to the distance between the tagged IAB node and the typical user. Obviously, the potential serving IAB donor nodes no longer the homogeneous PPP. Therefore, we consider the typical backhaul link with the assumption that the tagged IAB node is located at the origin for the sake of simplicity and emphasizes the analysis on the joint effects of resource partitioning and cache size allocation on the considered network performance.

are respectively given as

$$\mathcal{A}_{bm}^{\mathcal{L}} = \int_0^{\infty} \exp\left[-2\pi\lambda_m \hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) - 2\pi\lambda_m Z(R)\right] \times 2\pi\lambda_m Z'(R)dR, \tag{13}$$

$$\mathcal{A}_{bm}^{\mathcal{N}} = \int_0^{\infty} \exp\left[-2\pi\lambda_m Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right) - 2\pi\lambda_m \hat{Z}(R)\right] \times 2\pi\lambda_m \hat{Z}'(R)dR. \tag{14}$$

Proof: The proof is shown in the Appendix A. □

B. STATISTICAL DISTRIBUTION OF THE "SERVING DISTANCE"

Having derived the association probabilities, we now have the PDFs of the "serving distances" from the serving IAB node/IAB donor node to the typical user in LOS and NLOS access links, which are given as [28]

$$\hat{f}_{r_{x_s^{\mathcal{L}}}}(R) = \frac{1}{\mathcal{A}_{as}^{\mathcal{L}}} \exp\left[-2\pi\lambda_s\left(\hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) + Z\left(R\right)\right) - 2\pi\lambda_m\left(\hat{Z}\left(\left(K_B R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right) + Z\left(\left(K_B R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right)\right)\right] \times 2\pi\lambda_s Z'(R), \tag{15}$$

$$\hat{f}_{r_{x_s^{\mathcal{N}}}}(R) = \frac{1}{\mathcal{A}_{as}^{\mathcal{N}}} \exp\left[-2\pi\lambda_s\left(\hat{Z}\left(R\right) + Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right)\right) - 2\pi\lambda_m\left(\hat{Z}\left(\left(K_B R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right)\right] \times 2\pi\lambda_s \hat{Z}'(R), \tag{16}$$

$$\hat{f}_{r_{x_m^{\mathcal{L}}}}(R) = \frac{1}{\mathcal{A}_{am}^{\mathcal{L}}} \exp\left[-2\pi\lambda_m\left(\hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) + Z\left(R\right)\right) - 2\pi\lambda_s\left(\hat{Z}\left(\left(K_B^{-1} R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B^{-1} R^{\alpha_{\mathcal{L}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right)\right] \times 2\pi\lambda_m Z'(R), \tag{17}$$

$$\hat{f}_{r_{x_m^{\mathcal{N}}}}(R) = \frac{1}{\mathcal{A}_{am}^{\mathcal{N}}} \exp\left[-2\pi\lambda_m\left(\hat{Z}\left(R\right) + Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right)\right) - 2\pi\lambda_s\left(\hat{Z}\left(\left(K_B^{-1} R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{N}}}}\right) + Z\left(\left(K_B^{-1} R^{\alpha_{\mathcal{N}}}\right)^{\frac{1}{\alpha_{\mathcal{L}}}}\right)\right)\right] \times 2\pi\lambda_m \hat{Z}'(R). \tag{18}$$

Similarly, the PDFs of the "serving distances" from the serving IAB donor node to the tagged IAB node are given as

$$\hat{f}_{r_{y_m^{\mathcal{L}}}}(R) = \frac{1}{\mathcal{A}_{bm}^{\mathcal{L}}} \exp\left[-2\pi\lambda_m \hat{Z}\left(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}\right) - 2\pi\lambda_m Z(R)\right] 2\pi\lambda_m Z'(R), \tag{19}$$

$$\hat{f}_{r_{y_m^{\mathcal{N}}}}(R) = \frac{1}{\mathcal{A}_{bm}^{\mathcal{N}}} \exp\left[-2\pi\lambda_m Z\left(R^{\frac{\alpha_{\mathcal{N}}}{\alpha_{\mathcal{L}}}}\right) - 2\pi\lambda_m \hat{Z}(R)\right] 2\pi\lambda_m \hat{Z}'(R). \tag{20}$$

C. AVERAGE CELL LOADS

Denote the association cell of an IAB donor node or IAB node located at x as $\mathcal{C}(x)$. Since the blockage effects lead to the LOS and NLOS transmissions, the coverage area of IAB donor nodes and IAB nodes no longer form weighted Poisson Voronoi (PV) tessellation as mentioned above. Consequently, the exact characterization of the association cell area in the current setup is extremely difficult. However, based on the least biased path loss association scheme and the typical backhaul assumption whereby we ignore the spatial correlation between the access and backhaul, the tractable characterization of the mean cell load of the tagged node is computed by multiplying the user density (for access links) or IAB node density (for backhaul links) with the scaled mean association cell area that is assumed to be the same as that of the typical PV association area (*i.e.*, $\mathcal{C}_{a_t}(x_t)$ with $t \in \{m, s\}$ is referred to as the typical mean association areas of the node at x_t providing access links to users and $\mathcal{C}_b(y_m)$ is referred to as mean association areas of the IAB donor node at y_m providing backhaul links to IAB nodes.) [26]. However, since the performance trend will not much affected by a constant scaling factor introduced by the linear scaling approximation for computing the mean association cell area of the tagged node as long as the association-probability-dependent cell load is included. Therefore, this work further simplifies this cell load characterization by neglecting the bias factor and directly use the mean association cell area of the typical PV for both tagged and non-tagged nodes [27]. Furthermore, it is important to mention that each IAB node has both cache hit and cache miss users in access links, while each IAB donor node has only the cache miss users in backhaul links, depending on the cache size of the IAB node but independent of the LOS and NLOS transmissions. By using the same notation for the original PPP to represent its cell load, the following remark first gives the summary of the cell load for the general NC scenario, where both cache hit and cache miss cases are not included. Then, by modifying the general case, the cache-aided cell load is derived.

Remark 2: The mean cell load of IAB donor nodes/IAB nodes in access links and the mean cell load of the IAB donor nodes in backhaul links under full load assumption are given as below [27].

1) ***Access link****: The mean number of associated users of the serving IAB donor node/the tagged IAB node located at x_t in the access link is given as $\Phi_u(\mathcal{C}_{at}(x_t)) = \frac{\lambda_u \mathcal{A}_{at}}{\lambda_t}$ with $t \in \{m, s\}$, where $\mathcal{A}_{at} = \mathcal{A}_{at}^{\mathcal{L}} + \mathcal{A}_{at}^{\mathcal{N}}$.*

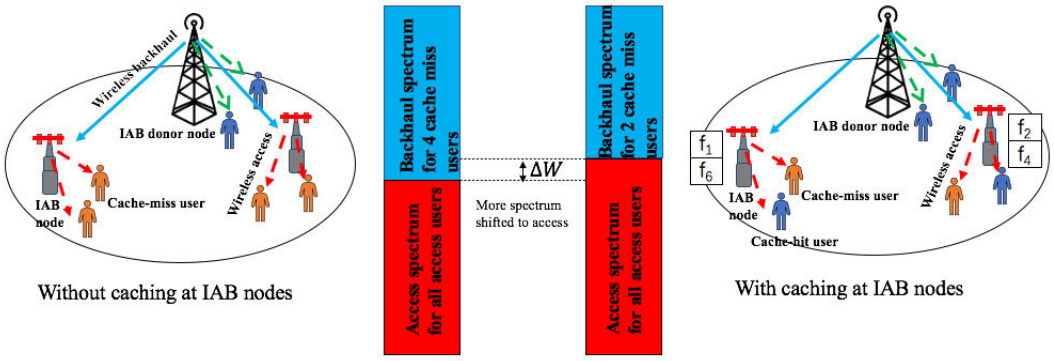

FIGURE 2. An illustration of spectrum resource partition policy for a cache-enabled IAB FR2 network.

2) ***Backhaul link***: *The mean number of IAB nodes of the serving IAB donor node located at y_m is given as $\Phi_s(\mathcal{C}_{bm}(y_m)) = \frac{\lambda_s}{\lambda_m}$. Subsequently, the total mean number of users and IAB nodes served by the serving IAB donor node on both access and backhaul is given as $\Phi_u(\mathcal{C}_{abm}(y_m)) = \Phi_u(\mathcal{C}_{am}(y_m)) + \Phi_s(\mathcal{C}_{bm}(y_m)) = \frac{\lambda_u \mathcal{A}_{am}}{\lambda_m} + \frac{\lambda_s}{\lambda_m}$.*

However, since each IAB node cannot serve more users than its available RF chains in one time-frequency resource block, we assume that the maximum number of users served by an IAB node in each resource block is limited to M_s^u such that $M_s^u \leq n_{\mathrm{RF}}^s$. Similarly, by assuming that half of the RF chains of each IAB donor node is used for the access and the other is used for the backhaul, each IAB donor node can serve total M_m^u users in access and M_m^s IAB nodes in backhaul such that $M_m^u \leq \frac{n_{\mathrm{RF}}^m}{2}$ and $M_m^s \leq \frac{n_{\mathrm{RF}}^m}{2}$, respectively. This work considers the best case where $M_s^u = n_{\mathrm{RF}}^s$, $M_m^u = \frac{n_{\mathrm{RF}}^m}{2}$, and $M_m^s = \frac{n_{\mathrm{RF}}^m}{2}$. Now let the set of all users that are scheduled and served in one resource block by the tagged IAB node located at x be denoted as $\mathcal{U}_x^s$. The cardinality of the set $\mathcal{U}_x^s$ is expressed as $U_x^s = \min(M_s^u, N_x^s)$, where $N_x^s = \frac{\lambda_u \mathcal{A}_{as}}{\lambda_s}$ is the number of associated users of the tagged IAB node given in the Remark 2. Similarly, let $\mathcal{U}_x^m$ denote the set of all users and IAB nodes served by the IAB donor node located at x such that its cardinality is given as $U_x^m = \min(M_m^u, N_x^{mu}) + \min(M_m^s, N_x^{ms})$, where $N_x^{mu} = \frac{\lambda_u \mathcal{A}_{am}}{\lambda_m}$ is the associated users in access and $N_x^{ms} = \frac{\lambda_s}{\lambda_m}$ is the associated IAB nodes in backhaul according to the Remark 2. Hereinafter, for notational simplicity, the average number of served users of the tagged IAB node is denoted as U_s while the average number of served users plus IAB nodes of the serving IAB donor node is denoted as U_m.

Further, for the caching scenario, the cache-enabled IAB nodes result in the traffic offloading on the backhaul link due to the cache hit event. Therefore, the IAB nodes require the backhaul link to serve the cache-miss users and hence, the load of backhaul link should be scaled by the factor $p_{\mathrm{miss}} = 1 - p_{\mathrm{hit}} = \sum_{i=1}^{F} q_i(1 - \omega_i)$, which will be given in details in the following section.

IV. RESOURCE ALLOCATION AND RATE CHARACTERIZATION

A. SPECTRUM RESOURCE ALLOCATION POLICY

The IAB framework allows both backhaul and access utilizing the same FR2 spectrum resource. More spectrum in backhaul can be save by implementing caching at IAB nodes as shown in Fig. 2. Therefore, the rate of access links is enhanced through the shifted spectrum resource. In such a cache-enabled FR2 IAB network, the total available bandwidth is assumed as W_{bw} Hz. However, since we do not apply full-duplex (FD) to the IAB node, the IAB node cannot simultaneously transmit in access and receive signals in backhaul at the same time. Subsequently, the spectrum resource allocated to backhaul cannot be utilized by the IAB node that is served by the backhaul to serve users in access. Besides, as mentioned in Section II-C, we consider OFDM-based hybrid beamforming transmissions at IAB donor nodes and IAB nodes, where the IAB donor node simultaneously serve users in access and the IAB nodes (*i.e.,* the cache miss users) on backhaul under the same spectrum resource at the same time through spatial multiplexing. Put another way, it means that the allocated spectrum resource to the backhaul is the same as the access of the IAB donor node. Thus, under this scenario the total bandwidth W_{bw} is divided into K subcarriers due to OFDM strategy and each orthogonal bandwidth per subcarrier is $\hat{W}_{bw} = \frac{W_{bw}}{K}$. In fact, all K subcarriers can be used for both IAB donor nodes and IAB nodes. However, due to the above discussion, we assume that we allocate N_{bw} subcarriers (*i.e.,* $0 \leq N_{bw} \leq K$) to the IAB donor node and subsequently, the rest $(K - N_{bw})$ subcarriers are allocated to its associated IAB node. Under this scheme in IAB framework, both IAB donor nodes and IAB nodes from other cells[7] can generate interference with each other. However, according to the literature [11], [29], [30], FR2 transmission tends to be the noise-limited scenario under dense networks with high blockage environment since the undesired signals are blocked and the interference is weak. Thus, next subsection we first give the received signal expression and then we ignore the

[7]Here we treat the IAB donor node and its served IAB nodes on backhaul and users in access as one cell.

interference term and characterize the SNR in this work to evaluate the performance. In particular, we also compare the special full cache case where all files are available in the local cache and thus, the backhaul is not required and all the bandwidth W_{bw} can be used for access by IAB nodes and IAB donor nodes at the same time.

B. SNR CHARACTERIZATION

1) ACCESS LINK [IAB DONOR NODE-/IAB NODE-USER LINK]

For access link, we consider the scenario where the typical user is associated with either the IAB donor node or IAB node. In particular, each IAB donor node (IAB node) simultaneously serves many scheduled users in access plus IAB nodes in backhaul (only users in access) through a fully-connected hybrid beamforming architecture. Let the BB and RF precoder matrices of the IAB donor node/IAB node located at x_t^j to serve the typical user at the kth subcarrier in LOS or NLOS path with $j \in \{\mathcal{L}, \mathcal{N}\}$ and $t \in \{m, s\}$ referring to as the IAB donor node and IAB node as $\mathbf{V}_{x_t^j}^{BB} = [\mathbf{v}_{x_t^j 0}^{BB}, \mathbf{v}_{x_t^j 1}^{BB}, \ldots, \mathbf{v}_{x_t^j U_t-1}^{BB}]$ and $\mathbf{V}_{x_t^j}^{RF} = [\mathbf{v}_{x_t^j 0}^{\mathrm{RF}}, \mathbf{v}_{x_t^j 1}^{RF}, \ldots, \mathbf{v}_{x_t^j U_t-1}^{\mathrm{RF}}]$, respectively.[8] Therefore, the received signal through the kth subcarrier at the typical user from the IAB node at x_s^j after passing through the BB and RF precoders and RF combiner at the typical user is given as

$$y_{0\,x_s^j}[k] = (\mathbf{w}_{0 x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{x_s^j 0}[k] \mathbf{V}_{x_s^j}^{\mathrm{RF}} \mathbf{v}_{x_s^j 0}^{\mathrm{BB}} s_{0 x_s^j} + I_{\mathrm{ICI}}^{as} + I_{\mathrm{OCI}}^{as} + n_0, \tag{21}$$

where $\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}}$ is the RF combiner of the typical user and $n_0 \sim CN(0, \sigma^2)$ denotes the noise power. ICI refers to as intra cell interference, while OCI refers to as out of cell interference. Thus, we have I_{ICI}^{as} and I_{OCI}^{as} are given as

$$I_{\mathrm{ICI}}^{as} = \sum_{\substack{v \in \mathcal{U}_{x_s^j}, \\ v \neq 0}} (\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{x_s^j 0}[k] \mathbf{V}_{x_s^j}^{\mathrm{RF}} \mathbf{v}_{x_s^j v}^{\mathrm{BB}} s_{v x_s^j}, \tag{22}$$

and

$$I_{\mathrm{OCI}}^{as} = \underbrace{\sum_{b \in \Phi_s, b \neq x_s^j} \sum_{u \in \mathcal{U}_b} (\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{b0}[k] \mathbf{V}_b^{\mathrm{RF}} \mathbf{v}_{bu}^{\mathrm{BB}} s_{ub}}_{\text{The interference from other IAB nodes except } x_s^j} + \underbrace{\sum_{\tilde{t} \in \Phi_m} \sum_{q \in \mathcal{U}_{\tilde{t}}} (\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{\tilde{t}0}[k] \mathbf{V}_{\tilde{t}}^{\mathrm{RF}} \mathbf{v}_{\tilde{t}q}^{\mathrm{BB}} \tilde{s}_{q\tilde{t}}}_{\text{The interference from the other IAB donor nodes}}. \tag{23}$$

In particular, since each subcarrier is assumed to be equally allocated with the transmit power $\frac{\mathrm{P}_t}{K}$ with $t \in \{m, s\}$, the transmitted symbol $s_{(*)}$ from the IAB node with the average power $\frac{\mathrm{P}_s}{KU_s}$. Similarly, the transmit symbol $\tilde{s}_{(*)}$ from the IAB donor node with the average power $\frac{\mathrm{P}_m}{KU_m}$. As for BB precoder, we consider zero forcing (ZF) strategy to eliminate the ICI such that $\mathbf{v}_{x_s^j}^{\mathrm{BB}} = \bar{\mathbf{h}}_{x_s^j}^H (\bar{\mathbf{h}}_{x_s^j 0}^{x_s^j})^H)^{-1}$, where $\bar{\mathbf{h}}_{x_s^j 0} = (\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{x_s^j 0}[k] \mathbf{V}_{x_s^j}^{\mathrm{RF}}$.

Similarly, based on (21), the received signal through the kth subcarrier at the typical user from the IAB donor node at x_m^j is given as

$$y_{0\,x_m^j}[k] = (\mathbf{w}_{0 x_m^j}^{\mathrm{RF}})^H \mathbf{H}_{x_m^j 0}[k] \mathbf{V}_{x_m^j}^{\mathrm{RF}} \mathbf{v}_{x_m^j 0}^{\mathrm{BB}} s_{0 x_m^j} + I_{\mathrm{ICI}}^{am} + I_{\mathrm{OCI}}^{am} + n_0, \tag{24}$$

where the ICI and OCI can be similarly given as shown in (22) and (23), respectively, which are omitted.

As mentioned that the interference in FR2 transmission can be ignored, next we give the SNR expressions at the typical user from the IAB node and IAB donor node, respectively. However, to derive the ASP of file delivery needs sum-rate over all allocated subcarriers, which is the function of k. Since the variable k only appears in the complex exponential function in $\omega_{\tau_l}[k]$ term, we consider the best-case scenario such that $\omega_{\tau_l} \leq \sum_{d=0}^{D-1} p_{\mathrm{rc}}(dT_s - \tau_i)$ so that it is independent of k. Subsequently, the precoders and combiners are independent of k as well. Further, since the analysis of this work is conducted from the probabilistic point of view, we consider each path experiences the same path delay $\tau_l = \tau = \frac{DT_s}{2}$ for $\forall l \in \{1, \ldots, \eta_{xy}\}$ with $\eta_{xy} \in \{\eta_{\mathcal{L}}, \eta_{\mathcal{N}}\}$ in an average sense. By which, we can simplify and avoid taking average with respect to delay over each path in the analysis. For notational simplicity, we can assume $\omega_{\tau_l} = \omega = \sum_{d=0}^{D-1} p_{\mathrm{rc}}(dT_s - \tau)$ as a constant.

As for the hybrid precoders, we consider the sub-optimal approach in [16] such that $\mathbf{w}_{0\,x_s^t}^{\mathrm{RF}} = \mathbf{a}_u(\theta_{l_{\max} x_t^j 0})$ and $\mathbf{v}_{x_t^j 0}^{\mathrm{RF}} = \mathbf{a}_t(\phi_{l_{\max} x_t^j 0})$ with $t \in \{m, s\}$. In particular, $\theta_{l_{\max} x_t^j 0}$, and $\phi_{l_{\max} x_t^j 0}$ are chosen such that the maximum channel gain is achieved on the $l_{\max} = \arg\max_l |h_{l x_t^j 0}|$. Hereinafter, we ignore the subscript $l_{\max}$ for notational simplicity.

Based on the received signal given in (21), the SNR of the typical user from the IAB node at x_s^j with $j \in \{\mathcal{L}, \mathcal{N}\}$ is formulated as shown below

$$\mathrm{SNR}_{as}^j[k] = \frac{\frac{\mathrm{P}_s}{U_s K} \left| (\mathbf{w}_{0\,x_s^j}^{\mathrm{RF}})^H \mathbf{H}_{x_s^j 0}[k] \mathbf{V}_{x_s^j}^{\mathrm{RF}} \mathbf{v}_{x_s^j 0}^{\mathrm{BB}} \right|^2}{\sigma^2}. \tag{25}$$

Since the magnitude of the normalized array factor is calculated as

$$\begin{aligned} G(\theta_1 - \theta_2) &= |\mathbf{a}^H(\theta_1)\mathbf{a}(\theta_2)|^2 \\ &= \frac{1}{n^2} \left| \sum_{i=0}^{n-1} e^{j2\pi i(\Theta_1 - \Theta_2)} \right|^2 = \frac{\sin^2[\pi n(\Theta_1 - \Theta_2)]}{n^2 \sin^2[\pi(\Theta_1 - \Theta_2)]}, \end{aligned} \tag{26}$$

where $\Theta_1 = \sin(\theta_1)$ and $\Theta_2 = \sin(\theta_2)$. Thus, when $\theta_1 = \theta_2$, $G(\theta_1 - \theta_2) = 1$; otherwise, it is approximated as zero due to the ON/OFF model [16]. Therefore, in order to acquire the

[8] Even though the BB precoders and combiners are related to the variable k, we do not use it in precoders' and combiners' notations for notational simplicity.

tractable SNR expression, we introduce the penalty term p_{PT} by considering the following assumptions [31]:

1) n_t^s and n_r^u are sufficiently large and
2) $n_t^s \gg U_s$.

Based on the suboptimal hybrid precoders and this ON/OFF assumption [16], the denominator in (25) for $l_{\max}$ path is approximated given as

$$\frac{\mathrm{P}_s}{U_s K}\left|(\mathbf{w}^{\mathrm{RF}}_{0\,x_s^j})^H \mathbf{H}_{x_s^j 0}[k]\mathbf{V}^{\mathrm{RF}}_{x_s^j}\mathbf{v}^{\mathrm{BB}}_{x_s^j 0}\right|^2$$
$$\approx \frac{\mathrm{P}_s p_{\mathrm{PT}}}{U_s K}\left|\mathbf{a}_u^H(\theta_{x_s^j 0})\sqrt{\frac{n_t^s n_r^u}{r_{x_s^j 0}^{\alpha_j}\eta_j}}h_{x_s^j 0}\omega \mathbf{a}_u^H(\theta_{x_s^j 0})\mathbf{a}_s(\phi_{x_s^j 0})\mathbf{a}_s(\phi_{x_s^j 0})\right|^2 \tag{27}$$
$$= \frac{\mathrm{P}_s}{U_s K}\frac{n_t^s n_r^u}{r_{x_s^j 0}^{\alpha_j}\eta_j}|h_{x_s^j 0}|^2|\omega|^2 p_{PT}, \tag{28}$$

where the penalty term is characterized as [31]

$$p_{\mathrm{PT}} = \begin{cases} 1 \text{ w.p. } (1-\frac{1}{n_t^s})^{(U_s-1)} \\ 0 \text{ otherwise} \end{cases}. \tag{29}$$

By using the above result, the tractable SNR can be reduced to

$$\mathrm{SNR}^j_{as}[k] \approx \frac{\frac{\mathrm{P}_s}{U_s K}\frac{n_t^s n_r^u}{\eta_j}|h_{x_s^j 0}|^2 r_{x_s^j 0}^{-\alpha_j}|\omega|^2 p_{\mathrm{PT}}}{\sigma^2}. \tag{30}$$

Similarly, the SNR at the typical user from the IAB donor node is given as

$$\mathrm{SNR}^j_{am}[k] = \frac{\frac{\mathrm{P}_m}{U_m K}\frac{n_t^m n_r^u}{\eta_j}|h_{x_m^j 0}|^2 r_{x_m^j 0}^{-\alpha_j}|\omega|^2 p_{\mathrm{PT}}}{\sigma^2}, \tag{31}$$

where $j \in \{\mathcal{L}, \mathcal{N}\}$.

2) BACKHAUL LINK [IAB DONOR NODE-IAB NODE LINK]

For typical backhaul link as mentioned above, we consider the scenario where the typical user is associated with the tagged IAB node but the requested file is not stored the cache. Therefore, the tagged IAB node should retrieve the non-cached file from the IAB donor node at y_m^j with $j \in \{\mathcal{L}, \mathcal{N}\}$ through the backhaul. In the similar manner, after passing through the analog combiner $\mathbf{w}^{\mathrm{RF}}_{0\,y_m^j}$ of the tagged IAB node, the wideband geometric channel model $\mathbf{H}_{y_m^j 0}$, and the BB and RF precoders of the IAB donor node, the received signal, where all transmitted signals to the cache miss users in the served IAB nodes are encoded together as one symbol, at the kth subcarrier at the tagged IAB node is given as

$$\bar{y}_{0x_m^j} = (\mathbf{w}^{\mathrm{RF}}_{0y_m^j})^H \mathbf{H}_{y_m^j 0}[k]\mathbf{V}^{\mathrm{RF}}_{y_m^j}\mathbf{v}^{\mathrm{BB}}_{y_m^j 0} + I^{bm}_{\mathrm{ICI}} + I^{bm}_{\mathrm{OCI}} + n_0, \tag{32}$$

where I^{bm}_{ICI} and I^{bm}_{OCI} can be given in a similar manner according to (22) and (23), respectively. Further, the SNR is approximately given as

$$\mathrm{SNR}^j_{bm}[k] = \frac{\frac{\mathrm{P}_m}{U_m K}\frac{n_t^m n_r^s}{\eta_j}|h_{y_m^j 0}|^2 r_{y_m^j 0}^{-\alpha_j}|\omega|^2 p_{\mathrm{PT}}}{\sigma^2}, \tag{33}$$

where $j \in \{\mathcal{L}, \mathcal{N}\}$.

C. RATE CHARACTERIZATION

Due to the aforesaid, we now characterize the rates of access and backhaul over allocated subcarriers from the caching perspective in this subsection. In the following, we characterize the rates from caching perspective in the 3 association events according to the spectrum resource allocation mentioned in Section IV-A.

- Case 1 (corresponding to $\mathcal{A}^{\mathcal{L}}_{am}$ and $\mathcal{A}^{\mathcal{N}}_{am}$): The typical user is associated with the IAB donor node with either LOS or NLOS. Therefore, the rate is given as
$$R^j_{am} = \sum_{k=1}^{N_{bw}} \frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{am}[k]), \tag{34}$$
where $j \in \{\mathcal{L}, \mathcal{N}\}$.
- Case 2 (corresponding to $\mathcal{A}^{\mathcal{L}}_{as}$, $\mathcal{A}^{\mathcal{N}}_{as}$, $\mathcal{A}^{\mathcal{L}}_{bm}$, and $\mathcal{A}^{\mathcal{N}}_{bm}$): The typical user is associated with the IAB donor node but the requested file is not stored. Therefore, the rates of the access and backhaul are given as
$$R^j_{as} = \sum_{k=1}^{K-N_{bw}} \frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{as}[k]), \tag{35}$$
$$R^j_{bm} = \sum_{k=1}^{N_{bw}} \frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{bm}[k]), \tag{36}$$
where $j \in \{\mathcal{L}, \mathcal{N}\}$.
- Case 3 (corresponding to $\mathcal{A}^{\mathcal{L}}_{as}$ and $\mathcal{A}^{\mathcal{N}}_{as}$): The typical user is associated with the IAB node but the requested file is stored. Therefore, the rate only in access is given as
$$R^j_{as} = \sum_{k=1}^{K-N_{bw}} \frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{as}[k]), \tag{37}$$
where $j \in \{\mathcal{L}, \mathcal{N}\}$.

Further, based on the aforementioned discussions and assumpitions, the rates are simplified as

- Case 1:
$$R^j_{am} = N_{bw}\frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{am}), \tag{38}$$
- Case 2:
$$R^j_{as} = (K-N_{bw})\frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{as}), \tag{39}$$
$$R^j_{bm} = N_{bw}\frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{bm}), \tag{40}$$
- Case 3:
$$R^j_{as} = (K-N_{bw})\frac{W_{bw}}{K}\log(1+\mathrm{SNR}^j_{as}), \tag{41}$$
where $j \in \{\mathcal{L}, \mathcal{N}\}$.

V. PERFORMANCE ANALYSIS

In this section we consider three key performance metrics, namely ASP, throughput, and backhaul latency of file delivery to evaluate the network performance of the considered system model. Before deriving these metrics, it is worth mentioning that in the traditional NC scenario, all the

users associated with the IAB node will be further served by the IAB donor nodes through the backhaul. However, under cache-enabled scenario, the cache-hit users associated with the IAB node will not require the backhaul. Therefore, the IAB node can acquire higher user experience for a given spectrum on backhaul link. In this work, we consider the average rate per link as the measurement criteria and assume the target rate per user is v_u. However, each backhaul link has many cache-miss users to serve. Therefore, the average target rate per link of the IAB donor node is calculated by

$$v_b = \frac{(X + Y \times Z)v_u}{\text{The total links of the IAB donor node}}, \tag{42}$$

where

$X =$ Number (No.) of associated access users of the IAB donor node,

$Y =$ No. of associated backhaul IAB nodes of the IAB donor node,

$Z =$ No. of cache miss users of those IAB nodes served by the IAB donor node.

And the total link is calculated by

$$\text{The no. of access users } + \text{The no. of backhaul IAB nodes.}$$

However, since each IAB node serves both cache hit and cache miss users, the average target rate per link is the same as the target rate per user v_u. When the cache size equals to the number of files, IAB donor nodes will not use backhaul links and $v_b = v_u$.

A. ASP OF FILE DELIVERY

The ASP of file delivery is defined as the probability of the instantaneous rate is greater than the pre-defined target data rate per link. With the association probabilities, the ASP of file delivery is computed by the summation of each conditional ASP of file delivery corresponding to the each association, which is given in the following Proposition.

Proposition 2: The ASP of file delivery under such a cache-enabled FR2 IAB network is given as

$$\begin{aligned}\mathcal{P}_s = \sum_{i=1}^{F} q_i \Big\{ & \mathcal{A}_{as}^{\mathcal{L}} \Big[\tilde{\omega}_i \mathbb{P}[R_{as}^{\mathcal{L}} \geq v_u] + (1-\tilde{\omega}_i)\mathbb{P}[R_{as}^{\mathcal{L}} \geq v_u] \\ & \times \Big(\mathcal{A}_{bm}^{\mathcal{L}} \mathbb{P}[R_{bm}^{\mathcal{L}} \geq v_b] + \mathcal{A}_{bm}^{\mathcal{N}} \mathbb{P}[R_{bm}^{\mathcal{N}} \geq v_b] \Big) \Big] \\ & + \mathcal{A}_{as}^{\mathcal{N}} \Big[\tilde{\omega}_i \mathbb{P}[R_{as}^{\mathcal{N}} \geq v_u] \\ & + (1-\tilde{\omega}_i)\mathbb{P}[R_{as}^{\mathcal{N}} \geq v_u] \Big(\mathcal{A}_{bm}^{\mathcal{L}} \mathbb{P}[R_{bm}^{\mathcal{L}} \geq v_b] \\ & + \mathcal{A}_{bm}^{\mathcal{N}} \mathbb{P}[R_{bm}^{\mathcal{N}} \geq v_b] \Big) \Big] \\ & + \Big[\mathcal{A}_{am}^{\mathcal{L}} \mathbb{P}[R_{am}^{\mathcal{L}} \geq v_b] + \mathcal{A}_{am}^{\mathcal{N}} \mathbb{P}[R_{am}^{\mathcal{N}} \geq v_b] \Big] \Big\}, \end{aligned} \tag{43}$$

where $v_b = \frac{(\frac{\lambda_u \mathcal{A}_{as}}{\lambda_m})p_{miss} + \frac{\lambda_u \mathcal{A}_{am}}{\lambda_m}}{\frac{\lambda_s}{\lambda_m} \times \tilde{p}_{miss} + \frac{\lambda_u \mathcal{A}_{am}}{\lambda_m}} v_u$ *and* $\tilde{p}_{miss} = 1 - [p_{hit}]^{U_s}$ *is the probability that there have cache miss users in each IAB node.*

Proof: The proof is given in Appendix B. □

B. AVERAGE THROUGHPUT

In order to evaluate the network capacity for the access and backhaul links, we define the average rate of file delivery of all links in the network for different caching strategies based on the following Proposition.

Proposition 3: The average throughput of file delivery under such a cache-enabled FR2 IAB network is given as

$$\begin{aligned}\bar{R} = & \sum_{i=1}^{F} q_i \Big\{ \mathcal{A}_{as}^{\mathcal{L}} \Big[\tilde{\omega}_i \bar{R}_a^{\mathcal{L}} + (1-\tilde{\omega}_i) \min(\bar{R}_a^{\mathcal{L}}, \mathcal{A}_{bm}^{\mathcal{L}} \bar{R}_{bm}^{\mathcal{L}} + \mathcal{A}_{bm}^{\mathcal{N}} \bar{R}_{bm}^{\mathcal{N}}) \Big] \\ & + \mathcal{A}_{as}^{\mathcal{N}} \Big[\tilde{\omega}_i \bar{R}_{as}^{\mathcal{N}} + (1-\tilde{\omega}_i) \min(\bar{R}_{as}^{\mathcal{N}}, \mathcal{A}_{bm}^{\mathcal{L}} \bar{R}_{bm}^{\mathcal{L}} + \mathcal{A}_{bm}^{\mathcal{N}} \bar{R}_{bm}^{\mathcal{N}}) \Big] \\ & + (\mathcal{A}_{am}^{\mathcal{L}} \bar{R}_{am}^{\mathcal{L}} + \mathcal{A}_{am}^{\mathcal{N}} \bar{R}_{am}^{\mathcal{N}}) \Big\}, \end{aligned} \tag{44}$$

where the rates in (44) *for the given target rate per user* τ_u *are given as*

$$\bar{R}_{as}^{j} = \mathbb{P}[R_{as}^{j} \geq v_u] v_u, \tag{45}$$

and considering the same target rate requirement as that of each user to every IAB node, we have the average rates for IAB donor nodes as given

$$\bar{R}_{tm}^{j} = \mathbb{P}[R_{tm}^{j} \geq v_b] v_u, \tag{46}$$

where $j \in \{\mathcal{L}, \mathcal{N}\}$.

C. AVERAGE LATENCY ON BACKHAUL LINK

Based on Proposition 3, the average latency is computed by

$$\begin{aligned}\bar{T} = & \sum_{i=1}^{F} q_i \Big\{ \mathcal{A}_{as}^{\mathcal{L}} \Big[\tilde{\omega}_i \frac{S}{\bar{R}_a^{\mathcal{L}}} + (1-\tilde{\omega}_i) \Big(\frac{S}{\bar{R}_a^{\mathcal{L}}} + \frac{S}{\mathcal{A}_{bm}^{\mathcal{L}} \bar{R}_{bm}^{\mathcal{L}} + \mathcal{A}_{bm}^{\mathcal{N}} \bar{R}_{bm}^{\mathcal{N}}} \Big) \Big] \\ & + \mathcal{A}_{as}^{\mathcal{N}} \Big[\tilde{\omega}_i \frac{S}{\bar{R}_{as}^{\mathcal{N}}} + (1-\tilde{\omega}_i) \Big(\frac{S}{\bar{R}_{as}^{\mathcal{N}}} + \frac{S}{\mathcal{A}_{bm}^{\mathcal{L}} \bar{R}_{bm}^{\mathcal{L}} + \mathcal{A}_{bm}^{\mathcal{N}} \bar{R}_{bm}^{\mathcal{N}}} \Big) \Big] \\ & + \mathcal{A}_{am}^{\mathcal{L}} \frac{S}{\bar{R}_{am}^{\mathcal{L}}} + \mathcal{A}_{am}^{\mathcal{N}} \frac{S}{\bar{R}_{am}^{\mathcal{N}}} \Big\}. \end{aligned} \tag{47}$$

Similarly, the average backhaul latency is given as

$$\bar{T}_b = \sum_{i=1}^{F} q_i (1-\tilde{\omega}_i)(\mathcal{A}_{as}^{\mathcal{L}} + \mathcal{A}_{as}^{\mathcal{N}}) \frac{S}{\mathcal{A}_{bm}^{\mathcal{L}} \bar{R}_{bm}^{\mathcal{L}} + \mathcal{A}_{bm}^{\mathcal{N}} \bar{R}_{bm}^{\mathcal{N}}}, \tag{48}$$

and the access latency is calculated by $\bar{T}_a = \bar{T} - \bar{T}_b$.

VI. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the WEC-aided FR2 IAB network with respect to the following metrics: ASP, throughput, and latency of file delivery under two commonly used caching strategies (*i.e.,* UC and MC) and compare them with the baseline NC scenario. In particular, we evaluate tradeoffs and the joint impacts of caching and spectrum resource allocation for different parameter settings on the performance of the considered

network model. Unless otherwise stated, the parameter settings used to produce the results are given as [9], [11], [32]: $\lambda_m = 10^{-5}$ nodes/m^2, $\lambda_s = 5 \times 10^{-5}$ nodes/m^2 nodes/m^2, $\lambda_u = 2.5 \times 10^{-4}$ nodes/m^2, $\mathrm{P}_m = 46$ dBm, $\mathrm{P}_s = 30$ dBm, $n_t^m = n_t^s = n_r^s = 256$, $n_r^u = 32$, $n_{\mathrm{RF}}^m = 40$, $n_{\mathrm{RF}}^s = 20$, $\alpha_{\mathcal{L}} = 2$, $\alpha_{\mathcal{N}} = 4$, $m_{\mathcal{L}} = 3$, $m_{\mathcal{N}} = 2$, $W_{\mathrm{bw}} = 1$ GHz, $\upsilon = 0.08$, $\upsilon_u = 5 \times 10^{-2} W_{\mathrm{bw}}$, $\beta = 0.02$, $D = 128$, $K = 512$, $S = 10$ Mbits.

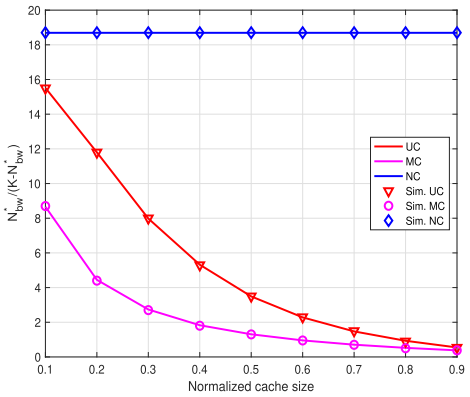

FIGURE 3. Resource transfer ratio v.s. normalized cache size for ASP of file delivery.

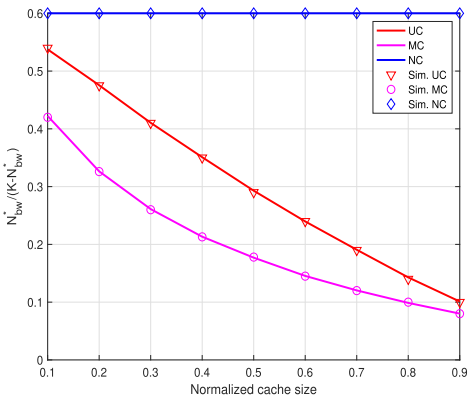

FIGURE 4. Resource transfer ratio v.s. normalized cache size for latency of file delivery.

1) Effect of caching on resource transfer ratio: First of all, we evaluate the resource transfer ratio, which is defined as the ratio of the optimal number of subcarriers allocated to the backhaul links (to achieve the best performance) and the remaining subcarriers allocated to the access link, *i.e.,* $\frac{N_{\mathrm{bw}}^*}{K - N_{\mathrm{bw}}^*}$. This is evaluated with respect to cache size for the ASP in Fig. 3, latency in Fig. 4, and throughput in Fig. 5. The results reveal that when the value of resource transfer ratio is greater than (less than) 1, more resource is allocated to backhaul (access), but in general a lower value of the resource transfer ratio means more savings in the backhaul bandwidth. Further, we can see from the figures that the resource transfer ratio decreases for both UC and MC starting from lower cache size to higher cache size. The reason behind this trend is

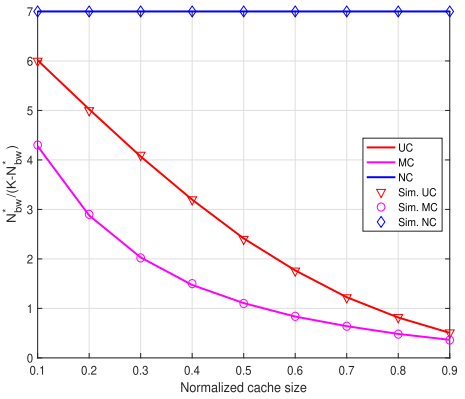

FIGURE 5. Resource transfer ratio v.s. normalized cache size for throughput of file delivery.

that when the cache size increases, more files are available in the local caches of the IAB nodes. Thus, less cache-miss users tend to less backhaul requirements and traffic, which lead to more spectrum resource shifted to the access for its performance enhancements. In particular, for the particular scenario where the cache size equals to the number of files, the resource transfer ratio becomes equal to be zero since there is no use of backhaul for IAB nodes to retrieve the files from the IAB donor nodes. Further, when compared to the NC scenario, it is seen that there are 87.7% (94.9%), 60% (75.8%), and 74.9% (88%) resource-transfer-ratio improvements (*i.e.,* relatively more backhaul savings) for the ASP, throughput, and latency of file delivery under UC (MC) for a normalized cache size of 0.6, respectively.

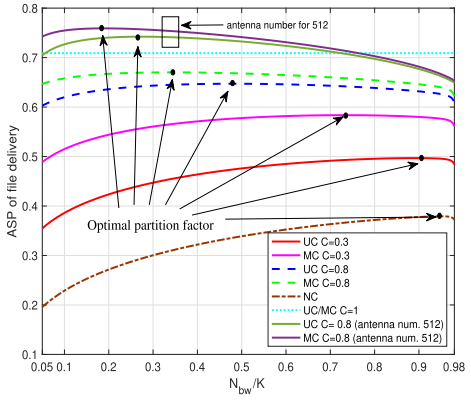

FIGURE 6. ASP of file delivery v.s. normalized partitioning factor with different cache size.

2) Effect of cache size and the number of antennas: In Fig. 6, 7, and 9, we respectively evaluate the significance of caching on the bandwidth resource allocation with respect to the ASP, latency, and throughput of file delivery for the default antenna settings as defined above. It can be seen from the figures that there exist an optimal resource partitioning factor. Similar to the discussion for Fig. 3, 4, and 5, as the cache size increases, the optimal resource partitioning factor decreases. However, for the case of average latency,

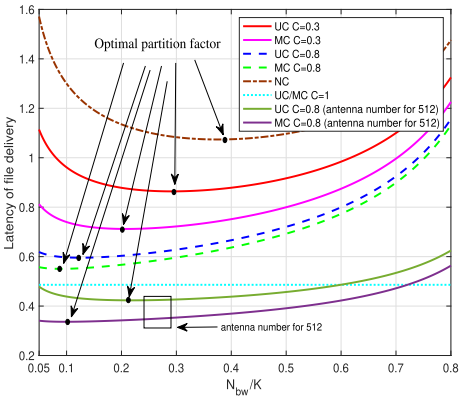

FIGURE 7. Latency of file delivery v.s. normalized partitioning factor with different cache size.

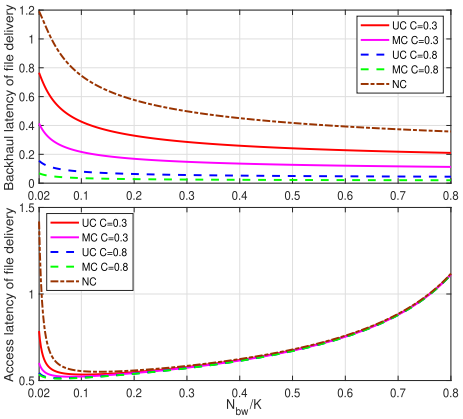

FIGURE 8. Backhaul/access latency of file delivery v.s. normalized partitioning factor with different cache size.

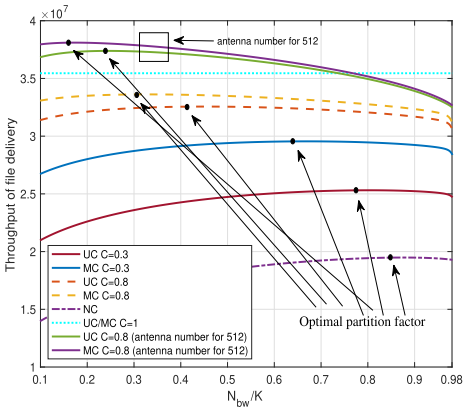

FIGURE 9. Throughput of file delivery v.s. normalized partitioning factor with different cache size.

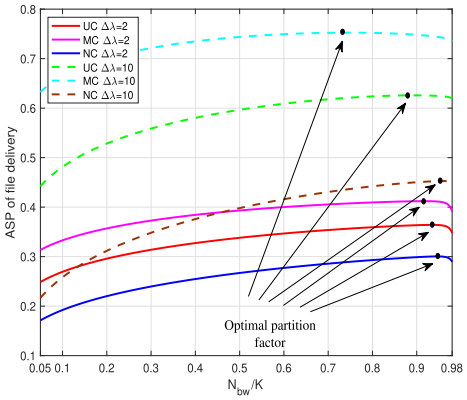

FIGURE 10. ASP of file delivery v.s. normalized partitioning factor with different IAB node density.

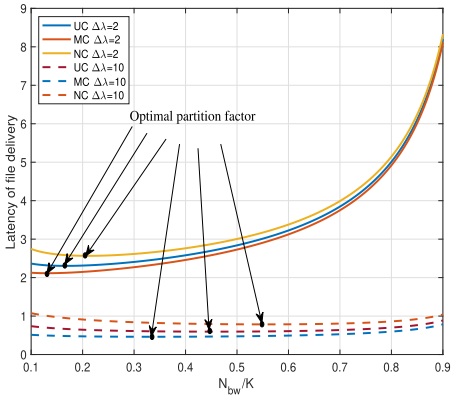

FIGURE 11. Latency of file delivery v.s. normalized partitioning factor with different IAB node density.

in Fig. 8 it can be seen that the performance trend for both access and backhaul is reversed with the increase in the resource partitioning factor. Further, for the sake of comparison in Fig. 9 we show the upper bound of the relevant performance for a special scenario (*i.e.,* full cache size), where the entire bandwidth is used in the access link. With increasing number of antennas, the performance is further improved to the extent that it betters even the upper bound with the same cache size if a proper resource partitioning factor is chosen. This is an innovative result since smaller size of antennas used for transmitting FR2 signals makes it practical to build very large arrays, thus compensating the performance loss that is incurred in the event of cache miss.

3) Effect of IAB node density: In Fig. 10, 11, and 12, we respectively evaluate the effect of IAB node density on the resource allocation for the normalized cache size of 0.3 with respect to the ASP, latency, and throughput of file delivery. The symbol $\Delta\lambda$ shown in the legend of the figures is defined as $\frac{\lambda_s}{\lambda_m}$. As shown in the figures, as the IAB node density increases, the performance of the network also improves. However, for the ASP and throughput of file delivery, the value of optimal resource partitioning factor decreases as the IAB node density increases, implying that more bandwidth in backhaul can be saved. On the contrary, optimal resource partitioning factor increases as the IAB node density increases with respect to the latency of file delivery. Therefore, there is a tradeoff between the ASP/throughput of file delivery and the latency of file delivery that design engineers must consider while selecting the optimal resource partitioning factor.

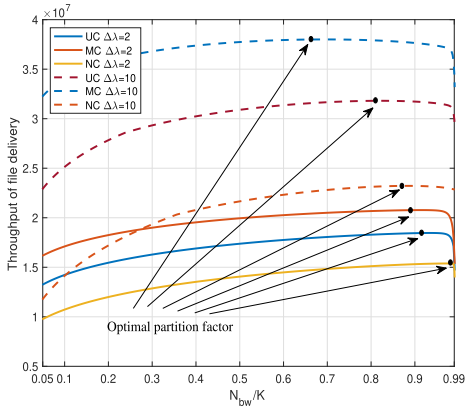

FIGURE 12. Throughput of file delivery v.s. normalized partitioning factor with different IAB node density.

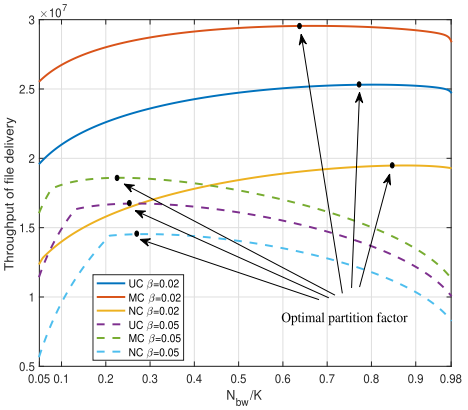

FIGURE 15. Throughput of file delivery v.s. normalized partitioning factor with different blockage density.

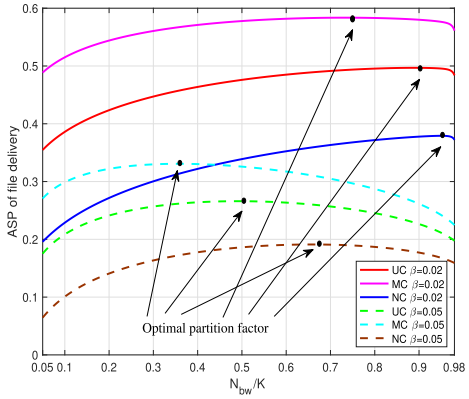

FIGURE 13. ASP of file delivery v.s. normalized partitioning factor with different blockage density.

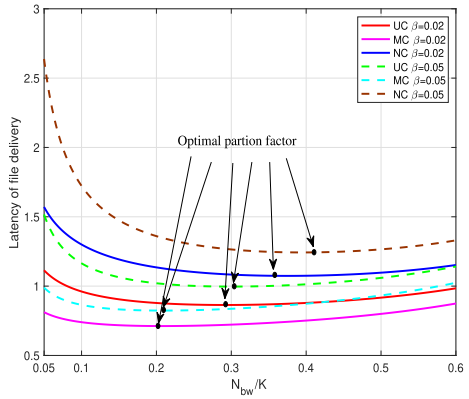

FIGURE 14. Latency of file delivery v.s. normalized partitioning factor with different blockage density.

4) Effect of blockage density: In Fig. 13, 14, and 15, we separately evaluate the effect of the blockage density on the bandwidth resource allocation with respect to the ASP, latency, throughput of file delivery for the normalized cache size of 0.3. In particular, when the blockage density increases, the performance of the network detoriates since there are more NLOS transmissions for both access and backhaul links. For the case of higher blockage density, the access should be allocated more spectrum resource than the backhaul to achieve optimal performance. This means that under heavy blockage scenario, the access link holds precedence over the backhaul. Hence, more storage is required to guarantee that the backhaul does not become a bottleneck of the network.

VII. CONCLUSION

The general concept of spectrum resource allocation in IAB networks using WEC with respect to the ASP, latency, and throughput of file delivery has been proposed in this paper to meet future trends in wireless networks. Comprehensive theoretical derivations as well as numerical results were provided to demonstrate that WEC is a cost-effective method for improving the IAB network's performance. With large storage capacity, sizeable chunks of backhaul bandwidth can be shifted towards the access for better network performance. Under various network parameter settings, it was observed that there always exists an optimal partitioning factor that achieves the optimal performance. In particular, it was seen that there is a tradeoff in the selection of the optimal partitioning factor with respect to ASP/ throughput and latency of file delivery for different IAB node densities. Our findings also revealed that in dense blockage environments, it is beneficial to give more resources to the access links. This means that higher cache size is necessary to ensure that the backhaul will not be the bottleneck of the IAB network.

The system model of this paper assumes that the users are uniformly distributed with specific QoS requirements. The problem formulation can be extended to take specific traffic patterns and constraints into account; non-uniform user distribution and heterogenous user preference can, for example, be used as a possible dimension to analyze. This is beyond the scope of this paper, but the analytical expressions in Propositions 1-3 can anyway be used to analyze a subset of the parameters with additional traffic constraints. Some other extensions are: i) to compare the energy consumption or energy efficiency of cache-enabled FR2 IAB networks with non-cached-enabled networks and ii) as pointed out

in [3], the cache content placement is a complex process and, the caching placement can be optimized if each user can connect to multiple IAB nodes. In both cases, the exact analysis would require a revised and more complicated system model and will be considered in the future.

.

APPENDIX A PROOF OF PROPOSITION 1

Let $r_{x_s^{\mathcal{L}}0}$, $r_{x_s^{\mathcal{N}}0}$, $r_{x_m^{\mathcal{L}}0}$, and $r_{x_m^{\mathcal{N}}0}$ be denoted as the distance from the LOS and NLOS IAB node (located at $x_s^{\mathcal{L}}$ and $x_s^{\mathcal{N}}$) as well as the LOS and NLOS IAB donor node (located at $x_m^{\mathcal{L}}$ and $x_m^{\mathcal{N}}$) to the typical user located at the origin. However, for notational simplicity, we ignore the subscript 0 throughout the proof. Based on the association event definitions, the even $E_{as}^{\mathcal{L}}$ means that the path loss seen at the typical user from the IAB node in LOS path is less than the other three events, namely $E_{as}^{\mathcal{L}} = \left\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_s r_{x_s^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}\right\} \bigcap \left\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}}\right\} \bigcap \left\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}\right\}$.

$$\begin{aligned}\mathcal{A}_{as}^{\mathcal{L}} &= \mathbb{E}_R\Big\{\mathbb{P}\Big[\Big\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_s r_{x_s^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}\Big\} \bigcap \Big\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}}\Big\} \\ &\quad \bigcap \Big\{B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}\Big\} \Big| r_{x_s^{\mathcal{L}}} = R\Big]\Big\} \\ &= \int_0^{\infty} \underbrace{\mathbb{P}[B_s R^{-\alpha_{\mathcal{L}}} > B_s r_{x_s^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}]}_{p_{ss}^{\mathcal{LN}}} \underbrace{\mathbb{P}[B_s R^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}}]}_{p_{sm}^{\mathcal{LL}}} \\ &\quad \underbrace{\mathbb{P}[B_s r_{x_s^{\mathcal{L}}}^{-\alpha_{\mathcal{L}}} > B_m r_{x_m^{\mathcal{N}}}^{-\alpha_{\mathcal{N}}}]}_{p_{sm}^{\mathcal{LN}}} f_{r_{x_s^{\mathcal{L}}}}(R) dR, \end{aligned} \tag{A.1}$$

where $f_{r_{x_s^{\mathcal{L}}}}(R)$ is the probability density function (PDF) of the nearest distance (also termed as "contact distance") between the typical user and the nearest IAB node. In the following, we derive $p_{ss}^{\mathcal{LN}}, p_{sm}^{\mathcal{LL}}, p_{sm}^{\mathcal{LN}}$, and $f_{r_{x_s^{\mathcal{L}}}}(R)$ in order. Based on the void probability [12], we have

$$\begin{aligned} p_{ss}^{\mathcal{LN}} &= \mathbb{P}[r_{x_s^{\mathcal{N}}} > R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}] \\ &= \mathbb{P}[\text{No IAB nodes from } \Phi_s^{\mathcal{N}} \text{ is closer than } R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}] \\ &= \exp\Big[-\int_0^{R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}} 2\pi\lambda_s p_{\mathcal{N}}(r_{x_s^{\mathcal{N}}}) r_{x_m^{\mathcal{N}}} dr_{x_s^{\mathcal{N}}}\Big] \\ &\overset{(a)}{=} \exp\Big[-\int_0^{R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}}} 2\pi\lambda_s (1-e^{-\beta r}) r dr\Big] \\ &\overset{(b)}{=} \exp\Big[-2\pi\lambda_s \hat{Z}(R^{\frac{\alpha_{\mathcal{L}}}{\alpha_{\mathcal{N}}}})\Big], \end{aligned} \tag{A.2}$$

where in (a) we change the variable $r_{x_s^{\mathcal{N}}}$ to r and substitute $1-e^{-\beta r}$ into $p_{\mathcal{N}}(r)$. In (b) we have $\hat{Z}(y) = \int_0^y (1-e^{-\beta r}) r dr = \frac{y^2}{2} + \frac{1}{\beta} y e^{-\beta y} + \frac{1}{\beta^2}(e^{-\beta y} - 1)$. Similarly, we have $p_{sm}^{\mathcal{LL}}$ and $p_{sm}^{\mathcal{LN}}$ that are given as

$$p_{sm}^{\mathcal{LL}} = \exp\Big[-2\pi\lambda_m Z((K_B R^{\alpha_{\mathcal{L}}})^{\frac{1}{\alpha_{\mathcal{L}}}})\Big], \tag{A.3}$$

where $K_B = \frac{B_m}{B_s}$ and $Z(y) = \int_0^y e^{-\beta r} r dr = -\frac{1}{\beta} y e^{-\beta y} - \frac{1}{\beta^2}(e^{-\beta y} - 1)$.

$$p_{sm}^{\mathcal{LN}} = \exp\Big[-2\pi\lambda_m \hat{Z}((K_B R^{\alpha_{\mathcal{L}}})^{\frac{1}{\alpha_{\mathcal{N}}}})\Big]. \tag{A.4}$$

Finally, the PDF of the "contact distance" $r_{x_s^{\mathcal{L}}}$ is derived by taking the first derivative over $\mathbb{P}[r_{x_s^{\mathcal{L}}} \le R]$. According to the void probability, $\mathbb{P}[r_{x_s^{\mathcal{L}}} \le R] = 1 - e^{-\Lambda_{x_m^{\mathcal{L}}}((0,R])}$, where the intensity measure $\Lambda_{x_s^{\mathcal{L}}}((0,R]) = 2\pi \int_0^{\infty} \lambda_s e^{-\beta r} r \mathbf{1}(r < R) dr = 2\pi\lambda_s Z(R)$. Therefore, we have

$$f_{r_{x_s^{\mathcal{L}}}}(R) = \frac{d\mathbb{P}[r_{x_s^{\mathcal{L}}} \le R]}{dR} = 2\pi\lambda_s Z'(R) e^{-2\pi\lambda_s Z(R)}, \tag{A.5}$$

where $Z'(R) = e^{-\beta R} R$. By substituting (A.2), (A.3), (A.4), and (A.5) into (A.1), we have the final association probability expression. In the similar way, we can derive the other association probabilities to complete the proof.

APPENDIX B PROOF OF PROPOSITION 2

For the IAB node-user access link, the ASP of file delivery is given as

$$\begin{aligned} &\mathbb{P}[R_{as}^j \ge \nu_u] \\ &= \mathbb{E}_{r_{x_s^j 0}}\Big\{\mathbb{P}\Big[(K - N_{bw})\frac{W_{bw}}{K}\log(1+\text{SNR}_{as}^j) \ge \tau_u\Big]\Big\} \\ &= \mathbb{E}_{r_{x_s^j 0}}\{\mathbb{P}[\text{SNR}_{as}^j \ge Q_{as}^j]\} \\ &= \mathbb{E}_{r_{x_s^j 0}}\Big\{\mathbb{P}\Big[\frac{\frac{\text{P}_s}{U_s K}\frac{n_t^s n_r^u}{\eta_j}|h_{x_s^j 0}|^2 r_{x_s^j 0}^{-\alpha_j}|\omega|^2 p_{\text{PT}}}{\sigma^2} \ge Q_{as}^j\Big]\Big\} \\ &\overset{(a)}{=} \mathbb{E}_{r_{x_s^j 0}}\Big\{\mathbb{P}\Big[|h_{x_s^j 0}|^2 \ge \underbrace{\frac{Q_{as}^j \sigma^2}{\frac{\text{P}_s}{U_s K}\frac{n_t^s n_r^u}{\eta_j}|\omega|^2}}_{\tilde{Q}_{as}^j} r_{x_s^j 0}^{\alpha_j}\Big]\Big\} \\ &\quad \times \Big(1 - \frac{1}{n_t^s}\Big)^{U_s - 1} \\ &\overset{(b)}{=} \int_0^{\infty} (1 - \frac{1}{n_t^s})^{U_s-1} \mathbb{P}\Big[|h_{r_{x_s^j 0}}|^2 \ge \tilde{Q}_{as}^j R^{\alpha_j}\Big] f_{r_{x_s^j 0}}(R) dR \\ &\overset{(c)}{=} \int_0^{\infty} (1 - \frac{1}{n_t^s})^{U_s-1} \sum_{\epsilon=0}^{\hat{m}_j - 1} \frac{(\tilde{Q}_{as}^j R^{\alpha_j} \hat{m}_j)^{\epsilon}}{\epsilon!} \\ &\quad \times \exp[-\hat{m}_j \tilde{Q}_{as}^j R^{\alpha_j}] \hat{f}_{r_{x_s^j 0}}(R) dR, \end{aligned} \tag{B.1}$$

where $Q_{as}^j = 2^{\frac{\nu_u K}{W_{\text{bw}}(K - N_{\text{bw}})}} - 1$ and $a)$ is derived for penalty term. $b)$ is derived by taking the average over the distance $r_{x_s^j 0}$ with the PDF of the serving distance $\hat{f}_{r_{x_s^j 0}}(R)$ given in (15) and (16) depending on $j \in \{\mathcal{L}, \mathcal{N}\}$. (c) follows from the cumulative distribution function (CDF) of the gamma distribution. Similarly, the ASP of file delivery for the IAB

donor node-user access link is given as

$$\mathbb{P}[R^j_{am} \geq \tau_b] = \int_0^{\infty} (1 - \frac{1}{n_t^m})^{U_m - 1} \sum_{\epsilon=0}^{m_j - 1} \frac{(\tilde{Q}^j_{am} R^{\alpha_j} \hat{m}_j)^{\epsilon}}{\epsilon!} \times \exp[-\hat{m}_j \tilde{Q}^j_{am} R^{\alpha_j}] \hat{f}_{r_{\hat{y}^j_m 0}}(R) dR, \tag{B.2}$$

where $\tilde{Q}^j_{am} = \frac{Q^j_{am} \sigma^2 \ U_m K \eta_j}{\mathrm{P}_m n_t^m n_r^u |\omega|^2}$ and $Q^j_{am} = 2^{\frac{K v_b}{N_{\mathrm{bw}} W_{\mathrm{bw}}}} - 1$. And the ASP of file delivery for the IAB donor node-IAB node backhaul link is given as

$$\mathbb{P}[R^j_{bm} \geq \tau_b] = \int_0^{\infty} (1 - \frac{1}{n_t^m})^{U_m - 1} \sum_{\epsilon=0}^{\hat{m}_j - 1} \frac{(\tilde{Q}^j_{bm} R^{\alpha_j} \hat{m}_j)^{\epsilon}}{\epsilon!} \times \exp[-\hat{m}_j \tilde{Q}^j_{bm} R^{\alpha_j}] \hat{f}_{r_{\hat{y}^j_m 0}}(R) dR, \tag{B.3}$$

where $\tilde{Q}^j_{bm} = \frac{Q^j_{bm} \sigma^2 \ U_m K \eta_j}{\mathrm{P}_m n_t^m n_r^s |\omega|^2}$ and $Q^j_{bm} = 2^{\frac{v_b K}{N_{\mathrm{bw}} W_{\mathrm{bw}}}} - 1$.

REFERENCES

[1] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in millimeter wave integrated access and backhaul for 5G," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8195–8210, Dec. 2018.

[2] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[3] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[4] J. Song, H. Song, and W. Choi, "Optimal content placement for wireless femto-caching network," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4433–4444, Jul. 2017.

[5] E. Bastug, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 649–653.

[6] S. Tamoor-Ul-Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-Aho, "Caching in wireless small cell networks: A storage-bandwidth tradeoff," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1175–1178, Jun. 2016.

[7] Y. Zhu, G. Zheng, K.-K. Wong, S. Jin, and S. Lambotharan, "Performance analysis of cache-enabled millimeter wave small cell networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6695–6699, Jul. 2018.

[8] L. Wang, K.-K. Wong, S. Lambotharan, A. Nallanathan, and M. Elkashlan, "Edge caching in dense heterogeneous cellular networks with massive MIMO-aided self-backhaul," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6360–6372, Sep. 2018.

[9] T. Zhang, S. Biswas, and T. Ratnarajah, "On the performance of cache-enabled hybrid wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1818–1834, Mar. 2021.

[10] A. Papazafeiropoulos and T. Ratnarajah, "Modeling and performance of uplink cache-enabled massive MIMO heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8136–8149, Dec. 2018.

[11] S. Biswas, T. Zhang, K. Singh, S. Vuppala, and T. Ratnarajah, "An analysis on caching placement for millimeter–micro-Wave hybrid networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1645–1662, Feb. 2019.

[12] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[13] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 167–203, 1st Quart., 2017.

[14] F. A. Khan, H. He, J. Xue, and T. Ratnarajah, "Performance analysis of cloud radio access networks with distributed multiple antenna remote radio heads," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4784–4799, Sep. 2015.

[15] G. Quer, I. Pappalardo, B. D. Rao, and M. Zorzi, "Proactive caching strategies in heterogeneous networks with device-to-device communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5270–5281, Aug. 2018.

[16] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.

[17] O. Y. Kolawole, S. Vuppala, and T. Ratnarajah, "Multiuser millimeter wave cloud radio access networks with hybrid precoding," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3661–3672, Dec. 2018.

[18] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6626–6637, Oct. 2016.

[19] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3401–3415, May 2017.

[20] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.

[21] N. Garg, M. Sellathurai, V. Bhatia, B. N. Bharath, and T. Ratnarajah, "Online content popularity prediction and learning in wireless edge caching," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1087–1100, Feb. 2020.

[22] N. Garg, M. Sellathurai, V. Bhatia, and T. Ratnarajah, "Function approximation based reinforcement learning for edge caching in massive MIMO networks," *IEEE Trans. Commun.*, early access, Dec. 28, 2020, doi: 10.1109/TCOMM.2020.3047658.

[23] A. Thornburg, T. Bai, and R. W. Heath, Jr., "Performance analysis of outdoor mmWave ad hoc networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 4065–4079, Aug. 2016.

[24] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[25] S. Park, A. Alkhateeb, and R. W. Heath, Jr., "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[26] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[27] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G: Performance analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2669–2684, Dec. 2019.

[28] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[29] T. Bai and R. W. Heath, Jr., "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[30] Y. Zhu, L. Wang, K.-K. Wong, and R. W. Heath, Jr., "Secure communications in millimeter wave ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3205–3217, May 2017.

[31] M. N. Kulkarni, A. Alkhateeb, and J. G. Andrews, "A tractable model for per user rate in multiuser millimeter wave cellular networks," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2015, pp. 328–332.

[32] T. Zhang, S. Biswas, and T. Ratnarajah, "An analysis on wireless edge caching in in-band full-duplex FR2-IAB networks," *IEEE Access*, vol. 8, pp. 164987–165002, 2020.

TONG ZHANG received the M.Sc. degree in signal processing and communications from The University of Edinburgh, Edinburgh, U.K., in 2016, where he is currently pursuing the Ph.D. degree with the Institute for Digital Communications. His main research interests include wireless edge caching, millimeter-wave communications, and stochastic geometry.

SUDIP BISWAS (Member, IEEE) received the Ph.D. degree in digital communications from The University of Edinburgh (UEDIN), U.K., in 2017. He has industrial experience with Tata Consultancy Services, India (Lucknow and Kolkata), where he held the position of an Assistant Systems Engineer from 2010 to 2012. He held the position of a Research Associate with the Institute of Digital Communications, UEDIN, from 2017 to 2019. He currently works with the Indian Institute of Information Technology Guwahati (IIITG) as an Assistant Professor with the Department of Electronics and Communications Engineering. He leads research on signal processing for wireless communications, with a particular focus on 5G's long-term evolution, including transceiver design for full-duplex radios, wireless edge caching, comms-radar co-existence, and reconfigurable intelligent surface assisted communication. He was an Organizer of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Edinburgh, U.K., in 2016, and has been involved in EU FP7 projects: remote radio heads and parasitic antenna arrays (HARP) and dynamic licensed shared access (ADEL), a DST UKIERI Project on wireless edge caching and an EPSRC Project on NoMA. He is currently leading the SERB Funded Project "Signal Processing for Co-Existence Between Radar and Future Communication Systems" as the PI.

THARMALINGAM RATNARAJAH (Senior Member, IEEE) was the Head of the Institute for Digital Communications from 2016 to 2018. He is currently with the Institute for Digital Communications, The University of Edinburgh, Edinburgh, U.K., as a Professor in digital communications and signal processing. He has published more than 400 publications in these areas and holds four U.S. patents. He has supervised 16 Ph.D. students and 21 postdoctoral research fellows and raised more than 11 million USD of research funding. He was the coordinator of the EU projects ADEL (3.7M) in the area of licensed shared access for 5G wireless networks, HARP (4.6M) in the area of highly distributed MIMO, as well as EU Future and Emerging Technologies projects HIATUS (3.6M) in the area of interference alignment and CROWN (3.4M) in the area of cognitive radio networks. His research interests include signal processing and information theoretic aspects of beyond 5G wireless networks, full-duplex radio, mmWave communications, random matrices theory, interference alignment, statistical and array signal processing, and quantum information theory. He is a Fellow of the Higher Education Academy (FHEA). He was the Technical Co-Chair of the 17th IEEE International workshop on Signal Processing Advances in Wireless Communications, Edinburgh, in July 2016. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2015 to 2017.